Homework 3 - Probability review

---

1) Conditional expectation of discrete random variables. Suppose $p(x, y, z)$, the joint pmf of the random variables (RVs) X, Y, and Z, is given by

$$p(1,1,1) = \frac{1}{8},\ p(1,1,2) = \frac{1}{8},\ p(1,2,1) = \frac{1}{16},\ p(1,2,2) = 0,$$
$$p(2,1,1) = \frac{1}{4},\ p(2,1,2) = \frac{3}{16},\ p(2,2,1) = 0,\ p(2,2,2) = \frac{1}{4}.$$

Evaluate $\mathbb{E}\left[X \,\middle|\, Y = 2\right]$ and $\mathbb{E}\left[X \,\middle|\, Y = 2, Z = 1\right]$.

2) Conditional expectation of continuous random variables. The joint pdf of X and Y is given by

$$f(x, y) = \frac{e^{-x/y}e^{-y}}{y}, \quad x > 0,\ y > 0.$$

Show that $\mathbb{E}\left[X|Y = y\right] = y$.

3) Expected time to match the initial outcome. Let $X_{\mathbb{N}} = X_0, X_1, \ldots, X_n, \ldots$ be an i.i.d. sequence of RVs with pmf

$$p(j) = \mathrm{P}\left[X_n = j\right],\ j = 1, \ldots, m, \quad \sum_{j=1}^{m} p(j) = 1.$$

Define

$$N = \min\{n > 0 : X_n = X_0\}$$

and show that $\mathbb{E}\left[N\right] = m$. (Hint: What is the conditional distribution of N given $X_0 = j$?)

4) The trapped miner. A miner is trapped in a mine containing three doors. The first door leads to a tunnel that takes him to safety after two hours of travel. The second door leads to a tunnel that returns him to the mine after three hours of travel. The third door leads to a tunnel that returns him to the mine after five hours. Assume the miner is at all times equally likely to choose any of the doors. Let N denote the total number of doors selected before the miner reaches safety. Also, let T_i denote the travel time corresponding to the $i-$th choice, $i \geq 1$. Finally, let X denote the time when the miner reaches safety. The goal of this exercise is to calculate the expected amount of time till the miner reaches safety, i.e., $\mathbb{E}\left[X\right]$. We solved this in class, and the idea is to follow a different approach here. To determine $\mathbb{E}\left[X\right]$, first give an identity that relates X to N and the T_i, $i \geq 1$. What are the expectations $\mathbb{E}\left[N\right]$ and $\mathbb{E}\left[T_N\right]$? Next, evaluate $\mathbb{E}\left[\sum_{i=1}^{N} T_i \,\middle|\, N = n\right]$ (Hint: here N and the T_i are not independent.) Use iterated expectations and the preceding results to obtain $\mathbb{E}\left[X\right]$.

5) Limit of a sequence of random variables. Suppose that $X_{\mathbb{N}} = X_1, X_2, \ldots, X_n, \ldots$ is an i.i.d. sequence of RVs, each with mean μ and variance σ^2. Derive a simple expression for

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left(X_{2i+1} + X_{2i}\right)^2$$

and provide justification for the existence of the limit.

6) The black box. Suppose we have a black box that generates realizations from distribution function $F(x)$ that has mean μ and variance σ^2. One "experiment" consists of drawing a long sequence of i.i.d.

realizations from the box. We will perform the experiment many times. In the $j-$th experiment, denote the $i-$th i.i.d. draw from the box by $X_{i,j}$. Define

$$\begin{aligned}
Y_{j,k} &= \frac{1}{k}\sum_{i=1}^{k} X_{i,j}, \\
Q_{j,k} &= \frac{kY_{j,k} - k\mu}{\sqrt{k\sigma^2}}, \\
U_{j,k}(x) &= \frac{1}{k}\sum_{i=1}^{k} \mathbb{I}\{X_{i,j} \le x\}, \quad x \in \mathbb{R}, \\
V_{l,k}(x) &= \frac{1}{l}\sum_{j=1}^{l} \mathbb{I}\{Q_{j,k} \le x\}, \quad x \in \mathbb{R}.
\end{aligned}$$

Suppose that m_1, m_2, n_1, n_2 are very large numbers with $m_2 \gg m_1$ and $n_2 \gg n_1$.

A) Will Y_{j,m_1} and Y_{j,m_2} be approximately equal?

B) Will Q_{j,m_1} and Q_{j,m_2} be approximately equal?

C) Will $U_{j,m_1}(x)$ and $U_{j,m_2}(x)$ be approximately equal?

D) Will $V_{m_1,n_1}(x)$ and $V_{m_2,n_2}(x)$ be approximately equal?

When possible, specify the approximate values of the quantities in parts A-D. Justify your answer. (The purpose of this exercise is to test your intuition, so there is no need for formal mathematical arguments – just clear explanations.)

7) Decision making. In this exercise you will examine a model for decision making. You are presented with a number of possible choices, some better than others. However it is unknown a priori how good options are going to be. All the information available is the quality of earlier options that were rejected, the quality of the option being offered and the number of options available. To ground discussion, say you are selling your car for which you are going to receive J offers – this is a somewhat unrealistic assumption but let us live with it for a while. Offers will be of different value, and if all J of them were presented upfront, the car would be sold to the highest bidder. Alas, potential buyers make their offers in a random order and if not accepted they withdraw their bid – presumably, they can find a different seller willing to accept their offer.

Denote the n-th offer rank as $X_n \in \{1, 2, \ldots J\}$. If $X_n = 1$ the best offer was made at time n, $X_n = 2$ implies that the second-best offer was made at time n and in general $X_n = i$ means that the i-th best offer was made at time n. Since offers are made randomly, all possible $J!$ rank orderings are equally likely and the probability of the i-th best offer being made at time n is then $1/J$.

Your strategy for deciding which offer to accept is the following:

- You do not accept any of the first K offers. You are not an earnest seller during this phase, you are just probing the market.
- You select the L-th best offer out of these K, which we denote as X_0. Note that $L \le K$
- After rejecting the first K offers, you chose the next following offer that exceeds X_0. That is, for $n \ge K + 1$ you choose the first $X_n < X_0$ (smaller means higher rank, thus better offer). Denote the accepted offer as X, i.e., $X = X_n$.
- If you reach the last offer J, you become desperate and accept offer X_J, i.e., $X = X_J$.

A) Simulate an individual experiment. Write a Matlab function that accepts as inputs the number of offers J, the number of rejected offers K and the selection constant L. The function returns the rank X of the accepted offer and the time n at which this offer was accepted. (Hint: check out function `randperm` in Matlab.)

B) Probability distribution of the rank of the selected offer. Start fixing the selection constant $L = 1$. Notice that this strategy simplifies to selecting the first offer that is better than the first K offers. Also, fix the number of offers to $J = 50$ and the number of rejected offers to $K = 30$. Using the function of part A, estimate the pmf of the rank of the selected offer X. That is, estimate $\mathrm{P}[X = j]$ for $j = 1, 2, \ldots, J$. Repeat the estimation of $\mathrm{P}[X = j]$ for $L = 2$ and $L = 5$. You should be pleasantly surprised that you do reasonably well with this decision strategy.

C) Probability of selecting the best offer. We are now interested in the probability of selecting the best offer, that is $\mathrm{P}[X = 1]$. Fix the number of offers to $J = 50$. Start considering $L = 1$, and vary the number of rejected offers K between 1 and $J - 1$. Estimate the probability $\mathrm{P}[X = 1]$ as a function of K. Repeat the estimation of $\mathrm{P}[X = 1]$ as a function of K for $L = 2$ and $L = 5$. You should be, again, pleasantly surprised to discover that the probability of selecting the best offer can be made quite high.

D) Probability of selecting the last offer ($L = 1$). Fix $L = 1$. The probability of selecting the last offer can be thought of as the probability that the decision policy fails because it drives you into a desperate decision. While we could use a simulation to estimate this probability it is possible, therefore preferable, to estimate this probability analytically. We want to find the probability $\mathrm{P}[X = X_J]$. When do we select the last offer? Well, we select $X = X_J$ if all offers are worst than X_0. But X_0 is the best offer (recall $L = 1$) among the first K ones. Thus, we select $X = X_J$ only if the best offer was one of the first K ones – since we end up setting $X_0 = 1$ and rejecting all offers until getting desperate – or if the last offer is the best and the second best offer was among the first K – since in this case we set $X_0 = 2$ and reject all offers until reaching $X_J = 1$. This chain of argument is enough to let you establish that

$$\mathrm{P}[X = X_J] = \frac{K}{J} + \frac{1}{J} \times \frac{K}{J-1} \; .$$

Explain.

E) Probability of selecting the best offer ($L = 1$). Fix $L = 1$. The probability of selecting the best offer, $\mathrm{P}[X = 1]$ can also be computed analytically when $L = 1$. Your goal here is to prove that

$$\mathrm{P}[X = 1] = \frac{K}{J} \sum_{n=K+1}^{J} \frac{1}{n-1} \; . \tag{1}$$

To help you in this endeavor we suggest that you first condition on $X_n = 1$, that is the probability of the n-th being the best possible offer. Then compute the conditional probability

$$\mathrm{P}\left[X = 1 \,\middle|\, X_n = 1\right] . \tag{2}$$

The above probability is the probability of selecting the best offer, given that the best offer is made at time n. Think about what this probability is when $n \leq K$ – i.e., when the best offer is made among the rejected ones – and when $n > K$. Use total probability to go from (2) to (1).

F) Optimal number of rejected offers K ($L = 1$). The expression in (2) is a function of K. It is possible to use it to find the optimal number of rejected offers K. For doing that we can use the approximation

$$\sum_{n=K+1}^{J} \frac{1}{n-1} \approx \int_{K}^{J-1} \frac{1}{x} dx \; . \tag{3}$$

Using (3) show that the optimal K is approximately J/e, where e is the base of natural logarithms (you can neglect that K is an integer in your maximization). Further show that for this value of K the probability of selecting the best offer is approximately $1/e \approx 0.37$. This might explain your pleasant surprises in parts B and C. Does it?